Bernoulli **19**(5A), 2013, 2098–2119
DOI: 10.3150/12-BEJ444

Marked empirical processes for non-stationary time series

NGAI HANG CHAN[1] and RONGMAO ZHANG[2]

[1]*Department of Statistics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong. E-mail: nhchan@sta.cuhk.edu.hk*
[2]*Department of Mathematics, Zhejiang University, Hangzhou, China. E-mail: rmzhang@zju.edu.cn*

Consider a first-order autoregressive process $X_i = \beta X_{i-1} + \varepsilon_i$, where $\varepsilon_i = G(\eta_i, \eta_{i-1}, \ldots)$ and $\eta_i, i \in \mathbb{Z}$ are i.i.d. random variables. Motivated by two important issues for the inference of this model, namely, the quantile inference for H_0: $\beta = 1$, and the goodness-of-fit for the unit root model, the notion of the marked empirical process $\alpha_n(x) = \frac{1}{n}\sum_{i=1}^{n} g(X_i/a_n) I(\varepsilon_i \le x), x \in \mathbb{R}$ is investigated in this paper. Herein, $g(\cdot)$ is a continuous function on $\mathbb{R}$ and $\{a_n\}$ is a sequence of self-normalizing constants. As the innovation $\{\varepsilon_i\}$ is usually not observable, the residual marked empirical process $\hat{\alpha}_n(x) = \frac{1}{n}\sum_{i=1}^{n} g(X_i/a_n) I(\hat{\varepsilon}_i \le x)$, $x \in \mathbb{R}$, is considered instead, where $\hat{\varepsilon}_i = X_i - \hat{\beta} X_{i-1}$ and $\hat{\beta}$ is a consistent estimate of β. In particular, via the martingale decomposition of stationary process and the stochastic integral result of Jakubowski (*Ann. Probab.* **24** (1996) 2141–2153), the limit distributions of $\alpha_n(x)$ and $\hat{\alpha}_n(x)$ are established when $\{\varepsilon_i\}$ is a short-memory process. Furthermore, by virtue of the results of Wu (*Bernoulli* **95** (2003) 809–831) and Ho and Hsing (*Ann. Statist.* **24** (1996) 992–1024) of empirical process and the integral result of Mikosch and Norvaiša (*Bernoulli* **6** (2000) 401–434) and Young (*Acta Math.* **67** (1936) 251–282), the limit distributions of $\alpha_n(x)$ and $\hat{\alpha}_n(x)$ are also derived when $\{\varepsilon_i\}$ is a long-memory process.

Keywords: goodness-of-fit; long-memory; marked empirical process; quantile regression; unit root

1. Introduction

Consider the autoregressive (AR) model

$$X_i = \beta X_{i-1} + \varepsilon_i, \tag{1.1}$$

where X_0 is given and $\varepsilon_i = G(\eta_i, \eta_{i-1}, \ldots,)$ is such that $\mathrm{E}\varepsilon_i = 0$ (when it exists) and $\{\eta_i\}$ is a sequence of i.i.d. random variables. It is known that when the tail of ε_i is heavy, the quantile estimate of β performs better than the least squares estimate (LSE). Even under the Gaussian setting, Zou and Yuan [39] proved that a composite quantile estimate can be as efficient as the maximum likelihood estimate (MLE). As a result, the quantile estimate provides a good alternate to the LSE. The first issue pursued in this paper is to study the asymptotic properties of the quantile estimate for model (1.1) with both long and short-memory innovations when $\beta = 1$.

A second motivation of this paper is to consider the goodness-of-fit issue for model (1.1). Empirical processes and goodness-of-fit tests in the i.i.d. case have long been a vibrant research topic in statistics, see, for example, the succinct monograph of del Barrio, Deheuvels and van de Geer [13], the proceeding of Gaenssler and Révész [17] and the references therein. Recently, there have also been developments on model checking using goodness-of-fit ideas for dependent data. For example, Bai [3] applied a Rosenblatt-transform to test the conditional distribution of ε_1 under condition on $\{\varepsilon_i, i \le 0\}$, Escanciano [14] and Hong and Lee [21] used a generalized spectral method to check the model fitness, Koul and Ling [26] considered the Kolmogorov–Smirnov (K–S) statistics of empirical process for GARCH model and Chan and Ling [7] generalized the K–S test to long-memory time series. It should be noted that Chan and Ling [7] only made use of the marginal distribution information of ε_1. For long-memory dependence, using only the marginal distribution information may reduce the test power and lead to incorrect conclusions. An alternative statistic which increases the power and takes into account of the dependent information is therefore required. Recently, some progresses have been made on this issue. For example, Woodridge [35] and Escanciano [15] proposed the statistic $\sum_{i=1}^n g(X_{i-1})\varepsilon_i$ for a measurable weighted function $g(\cdot)$. Stute, Xu and Zhu [32] used $\sum_{i=1}^n g(X_{i-1})(I(\varepsilon_i \le x) - F(x))$ to test the validity of a model for independent data. The idea of Stute, Xu and Zhu was used by Escanciano [16] to check the joint specification of conditional mean and variance of a GARCH-type model.

It turns out that the key idea of studying these two issues lies in analyzing the asymptotic property of

$$\sum_{i=1}^n g(X_{i-1})\bigl[I(\varepsilon_i < x) - F(x)\bigr], \qquad x \in \mathbb{R}, \tag{1.2}$$

where $g(\cdot)$ is a measurable weighted continuous function on $\mathbb{R}$. Note that if X_n is a unit root model, then under some regularity conditions, there exists a constant sequence $\{a_n\}$ such that $X_{[nt]}/a_n \stackrel{\text{f.d.d.}}{\longrightarrow} \xi(t)$ for some random process $\xi(t)$, where $\stackrel{\text{f.d.d.}}{\longrightarrow}$ denotes the weak convergence of finite-dimensional distributions. This leads us to replace the statistic of Stute, Xu and Zhu [32] by

$$\alpha_n(x) = \sum_{i=1}^n g(X_{i-1}/a_n)\bigl(I(\varepsilon_i \le x) - F(x)\bigr). \tag{1.3}$$

Observe that $\alpha_n(x)$ is a general form of (1.2) and its limit behavior offers important insight in studying the two aforementioned issues.

Specifically, let $\{\eta_i'\}$ be an i.i.d. copy of $\{\eta_i\}$, $\mathcal{F}_i$ be the σ-field generated by $\{\eta_t, t \le i\}$, that is, $\mathcal{F}_i = \sigma(\eta_i, \eta_{i-1}, \ldots)$ and $\mathcal{F}_i^* = \sigma(\eta_i', \eta_{i-1}, \ldots, \eta_1, \eta_0, \eta_{-1}, \eta_{-2} \ldots,)$. Let L^p be the space of random variables Z with $\|Z\|_p = (\mathrm{E}|Z|^p)^{1/p} < \infty$. For simplicity, we also write $\|\cdot\|_2$ as $\|\cdot\|$. For $j \in \mathbb{Z}$, define the projection operator

$$\mathcal{P}_j \cdot = \mathrm{E}(\cdot|\mathcal{F}_j) - \mathrm{E}(\cdot|\mathcal{F}_{j-1})$$

and define the predictive dependence measure $\theta_p(i) = \|\mathcal{P}_0 \varepsilon_i\|_p$ as in Wu [37]. We say that a process $\{\varepsilon_i\}$ is a short-memory process if $\sum_{i=0}^\infty \theta_p(i) < \infty$, otherwise, we say it is a long-memory process.

The main purpose of this paper is to consider a unified approach for the limit of (1.3) and the statistic

$$\hat{\alpha}_n(x) = \sum_{i=1}^{n} g(X_{i-1}/a_n)\big(I(\hat{\varepsilon}_i \le x) - F(x)\big) \tag{1.4}$$

for model (1.1) under non-stationarity with long and/or short-memory innovations $\{\varepsilon_i\}$, where $\hat{\varepsilon}_i = X_i - \hat{\beta} X_{i-1}$ and $\hat{\beta}$ is an estimate of β. Although Escanciano [16] has also considered the model checking problem for dependent data (GARCH), his underlying model still possesses a martingale structure. When $\{X_i\}$ is a stationary process adapted to the fields $\mathcal{F}_i = \sigma(\varepsilon_i, \varepsilon_{i-1}, \ldots) = \sigma(\eta_i, \eta_{i-1}, \ldots)$ and $\{\varepsilon_i\}$ is a martingale difference sequence, the central limit theorem (CLT) for martingale differences can be applied to derive the limit distribution of (1.3). When X_i is a random walk process, it is not clear how to derive the limit distribution of (1.3), especially, if $\{\varepsilon_i\}$ is a long-memory process. This is because the CLT of martingale differences cannot be directly used and when $\{\varepsilon_i\}$ is a long-memory process, neither $g(X_{[nt]}/a_n)$ nor $\sum_{i=1}^{[nt]}(I(\hat{\varepsilon}_i < x) - F(x))$ can be approximated by a martingale. In this paper, we first use the result of Jakubowski [22] (see also Protter and Kurtz [27]) on the weak convergence of the stochastic integral to deduce the limit distributions of (1.3) and (1.4), when $\{\varepsilon_i\}$ is a short-memory process. We then combine the results of Wu [36] and Ho and Hsing [20] on the expression of empirical process and the integral result of Mikosch and Norvaiša [30] and Young [38] to deduce the limit distributions of $\alpha_n(x)$ and $\hat{\alpha}_n(x)$ when $\{\varepsilon_i\}$ is a long-memory process.

The paper is organized as follows. In Section 2, we consider the marked empirical process when $\{\varepsilon_i\}$ is short-memory. Section 3 considers the case with long-memory innovations. Proofs are given in Section 4.

2. Short-memory error processes

In this section, we consider the limit distribution for (1.3) and (1.4) when $\{\varepsilon_i = G(\eta_i, \eta_{i-1}, \ldots\}$ is a short-memory process with mean zero. Let $S_{[nt]} =: S_n(t) = \sum_{i=1}^{[nt]} \varepsilon_i$. According to Theorem 2* of Chapter 7 (see pages 162 and 175) of Gnedenko and Kolmogorov [18], if $\{\varepsilon_i\}$ are i.i.d. and there exists a sequence $\{a_n\}$ such that $S_n(1)/a_n \xrightarrow{\mathcal{L}} S(1)$, then $S(1)$ is a stable variable, where $\xrightarrow{\mathcal{L}}$ denotes the convergence in distribution. Further, when ε_i has an infinite variance, it must satisfy for any $y > 0$,

$$\lim_{x\to\infty} P\big(|\varepsilon_1| \ge xy\big)/P\big(|\varepsilon_1| \ge x\big) = y^{-\alpha} \tag{2.1}$$

and the normalization constants $\{a_n\}$ are given by

$$a_n = \inf\big\{x\colon\ P\big(|\varepsilon_1| \ge x\big) \le 1/n\big\}. \tag{2.2}$$

Similar behaviors exist for short-memory processes under certain regularity conditions, see, for example, Davis and Resnick [11]. Throughout the paper, we assume (2.1) and (2.2) hold when $\{\varepsilon_i\}$ has an infinite variance and there exists a constant sequence $\{a_n\}$ such that $S_n(1)/a_n \xrightarrow{\mathcal{L}} S(1)$.

Let $F_i(x|\mathcal{F}_j) = P(\varepsilon_i \le x|\mathcal{F}_j)$, $f_i(x|\mathcal{F}_j) = F_i^{(1)}(x|\mathcal{F}_j)$ be the conditional distribution (resp., density) function of ε_i at x given $\mathcal{F}_j$ and f_i be the marginal density of ε_i. Let $W_n(t,x) = \sum_{i=1}^{[nt]}[I(\varepsilon_i < x) - F(x)]$ and $W(t,x)$ be a rescaled Brownian bridge for fix t and a Brownian motion with variance $\mu(x) = E\{\sum_{i=0}^{\infty} F_i(x|\mathcal{F}_0) - F_i(x|\mathcal{F}_0^*)\}^2$ for fix x. Further, let $\stackrel{S}{\Longrightarrow}$ and $\stackrel{w}{\Longrightarrow}$ denote the weak convergence in S and J_1-topology, respectively, and impose the following assumptions:

(A1) $S_n(t)/a_n \stackrel{S}{\Longrightarrow} S(t)$ on $D[0,1]$. For more information about the weak convergence in the S-topology, see Jakobowski ([22] and [23]),

(A2) $g(\cdot)$ is a Hölder continuous function on $\mathbb{R}$, that is, $|g(x) - g(y)| \le C|x-y|^\nu$ for all $x, y \in (-\infty, \infty)$, where $\nu = 1$, when ε_1 has infinite variance with tail index $\alpha < 2$ and $\nu > 1$ when $\alpha = 2$ or ε_1 has finite variance.

(A3)

- (i) $\sum_{j=1}^{\infty} \| \sum_{i=j}^{\infty} F_i(x|\mathcal{F}_0) - F_i(x|\mathcal{F}_0^*)\|^2 < \infty$, or
- (ii) $\sum_{i=1}^{\infty} \|F_i(x|\mathcal{F}_0) - F_i(x|\mathcal{F}_0^*)\| < \infty$ and $\sum_{i=m}^{\infty} \|F_i(x|\mathcal{F}_0) - F_i(x|\mathcal{F}_0^*)\| = \mathrm{O}[(\log m)^{-a}]$, $a > 3/2$, when $\sum_{i=1}^{n} |\varepsilon_i|/a_n = \mathrm{O}_p(1)$,

(A4) $\sum_{j=1}^{\infty} \sup_x \| \sum_{i=j}^{\infty} F_i^{(l)}(x|\mathcal{F}_0) - F_i^{(l)}(x|\mathcal{F}_0^*)\|^2 < \infty, l = 0, 1, F_i^{(0)}(x|\mathcal{F}_0) = F_i(x|\mathcal{F}_0)$.

Theorem 2.1. *Suppose that conditions* (A1)–(A3) *hold*, *then for any* $x \in \mathbb{R}$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(S_{i-1}/a_n)\bigl[I(\varepsilon_i < x) - F(x)\bigr] \stackrel{\mathcal{L}}{\longrightarrow} \int_0^1 g\bigl(S(t-)\bigr)\,\mathrm{d}W(t,x). \tag{2.3}$$

In addition, *if* (A3) *is replaced by* (A4), *then for any constant* $A > 0$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(S_{i-1}/a_n)\bigl[I(\varepsilon_i < x) - F(x)\bigr] \stackrel{w}{\Longrightarrow} \int_0^1 g\bigl(S(t-)\bigr)\,\mathrm{d}W(t,x), \qquad \textit{on } D[-A,A], \tag{2.4}$$

Theorem 2.2. *Suppose that* $\beta = 1$ *in model* (1.1) *and conditions* (A1)–(A3) *in Theorem* 2.1 *hold*, *then*

$$\frac{1}{\sqrt{n}}\alpha_n(x) \stackrel{\mathcal{L}}{\longrightarrow} \int_0^1 g\bigl(S(t-)\bigr)\,\mathrm{d}W(t,x). \tag{2.5}$$

In addition, *if* $a_n(\hat{\beta} - 1) = \mathrm{o}_p(1)$, *then*

$$\begin{aligned}\frac{1}{\sqrt{n}}\hat{\alpha}_n(x) &= \frac{1}{\sqrt{n}}\alpha_n(x) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_{i-1}/a_n)\bigl[F\bigl(x + (\hat{\beta} - \beta)X_{i-1}\bigr) - F(x)\bigr] + \mathrm{o}_p(1).\end{aligned} \tag{2.6}$$

Let $\hat{\beta}$ be the τ-quantile estimate of β when ε_1 has infinite variance with tail index $\alpha < 2$, that is, $\hat{\beta} = \arg\min_\beta \sum_{i=1}^n \rho_\tau(X_i - \beta X_{i-1} - F^{-1}(\tau))$, where $\rho_\tau(y) = y(\tau - I(y \le 0))$. When $\beta = 1$, using the argument of Theorem 4 in Knight (see also Chan and Zhang [9]), we have

$$a_n\sqrt{n}(\hat{\beta} - \beta) = \frac{(1/\sqrt{n})\sum_{t=1}^n (X_{t-1}/a_n)(\tau - I(\varepsilon_t \le F^{-1}(\tau)))}{(1/n)\sum_{t=1}^n f_t(F^{-1}(\tau)|\mathcal{F}_{t-1})(X_{t-1}^2/a_n^2)} + \mathrm{o}_p(1).$$

By virtue of Theorem 2.2 and this expression, the following corollary concerning the quantile estimate is immediate.

Corollary 2.1. *Under conditions Theorem 2.2, if* $\mathrm{E}|f_1(F^{-1}(\tau))|\mathcal{F}_0|^p < \infty$ *for some* $p > 1$ *and* $f(F^{-1}(\tau)) > 0$, *then*

$$a_n\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} -\frac{1}{f(F^{-1}(\tau))} \frac{\int_0^1 S(t-)\,\mathrm{d}W(t, F^{-1}(\tau))}{\int_0^1 S^2(t)\,\mathrm{d}t}.$$

Theorem 2.3. *In addition to the conditions of Theorem 2.2, if* (A4) *holds, then for any constant* $A > 0$,

$$\sup_{x\in[-A,A]} \frac{1}{\sqrt{n}}\alpha_n(x) \xrightarrow{\mathcal{L}} \sup_{x\in[-A,A]} \int_0^1 g\bigl(S(t)\bigr)\,\mathrm{d}W(t,x). \tag{2.7}$$

For $\hat{\alpha}_n(x)$, *we have*:

(a) *if* $\hat{\beta}$ *is the* τ*-quantile estimate of* β *and* $f(F^{-1}(\tau)) > 0$, *then*

$$\begin{aligned}&\sup_{x\in[-A,A]} \frac{1}{\sqrt{n}}\hat{\alpha}_n(x)\\ &\quad\xrightarrow{\mathcal{L}} \sup_{x\in[-A,A]}\Biggl[\biggl(-\frac{f(x)}{f(F^{-1}(\tau))}\biggr)\biggl(\frac{\int_0^1 S(t-)\,\mathrm{d}W(t,F^{-1}(\tau))}{\int_0^1 S^2(t)\,\mathrm{d}t}\biggr)\\ &\qquad\times\int_0^1 g\bigl(S(t)\bigr)S(t)\,\mathrm{d}t + \int_0^1 g\bigl(S(t-)\bigr)\,\mathrm{d}W(t,x)\Biggr].\end{aligned} \tag{2.8}$$

(b) *if* $\hat{\beta}$ *is the LSE of* β *and* $(S_n(t), \sum_{i=1}^n \varepsilon_i^2/a_n^2) \xrightarrow{f.d.d.} (S(t), S^2)$, *then*

(i) *if* $a_n = n^\vartheta l(n)$ *for some* $1/2 < \vartheta < 1$,

$$\begin{aligned}&\sup_{x\in[-A,A]} \frac{1}{a_n}\hat{\alpha}_n(x)\\ &\quad\xrightarrow{\mathcal{L}} \sup_{x\in[-A,A]} f(x)\int_0^1 S(t-)\,\mathrm{d}S(t)\\ &\qquad\times\int_0^1 g\bigl(S(t)\bigr)S(t)\,\mathrm{d}t\Big/\int_0^1 S^2(t)\,\mathrm{d}t.\end{aligned} \tag{2.9}$$

(ii) *If* $a_n = \sqrt{n}$,

$$\sup_{x\in[-A,A]} \frac{1}{\sqrt{n}}\hat{\alpha}_n(x) \xrightarrow{\mathcal{L}} \sup_{x\in[-A,A]} \biggl[f(x)\int_0^1 S(t-)\,\mathrm{d}S(t)\int_0^1 g\bigl(S(t)\bigr)S(t)\,\mathrm{d}t \Big/ \int_0^1 S^2(t)\,\mathrm{d}t + \int_0^1 g\bigl(S(t-)\bigr)\,\mathrm{d}W(t,x)\biggr]. \tag{2.10}$$

Remark 2.1. By Volný [34], condition (i) in (A3) is a necessary and sufficient condition for $I(\varepsilon_i \le x)$ enjoying a martingale decomposition, that is, there exist a martingale difference $\zeta_i(x)$ with respect to $\mathcal{F}_i$ and a sequence $\xi_i(x) \in L^2, i \in \mathcal{Z}$ such that

$$I(\varepsilon_i \le x) - F(x) = \zeta_i(x) + \xi_i(x) - \xi_{i+1}(x). \tag{2.11}$$

Remark 2.2. From Theorems 2.2 and 2.3, we see that the limit distribution of the test statistics based on $\alpha_n(x)$ and $\hat{\alpha}_n(x)$ are very different in the unit root case. As a result, using a residual marked empirical process ($\hat{\alpha}_n(x)$) to test the goodness-of-fit of nonstationary processes will be very different from using the marked empirical process ($\alpha_n(x)$).

To illustrate the usefulness of these theorems, consider the following examples, which characterize the limit distributions of the marked empirical process α under various situations.

Example 2.1. Let $\{\varepsilon_i\}$ in model (1.1) be the generalized autoregressive conditional heteroscedasticity (GARCH(1, 1)) process

$$\varepsilon_i = \sigma_i\eta_i, \qquad \sigma_i^2 = \omega + a\sigma_{i-1}^2 + b\varepsilon_{i-1}^2,$$

where $\omega, a, b > 0$, $\{\eta_i\}$ is an i.i.d. symmetric random sequence with $\mathrm{E}[\log(a + b\eta_1^2)] < 0$ and $\mathrm{E}(a + b\eta_1^2)^r < \infty$ for some $r > 0$. If there exists a positive constant C_0 such that the density $f_\eta(\cdot)$ of η_1 satisfies $\sup_x f_\eta(x) < C_0$, according to Kesten [24] (see also Lemma A.1 in Chan and Zhang [10]), there exists an $\alpha > 0$ such that $\mathrm{E}(a + b\eta_1^2)^{\alpha/2} = 1$ and there exists a constant c_0 such that $\lim_{x\to\infty} x^{-\alpha}P(|\varepsilon_1| > x) = c_0$. Let η_0' be a independent copy of η_0, then there exist a constant C such that

$$\begin{aligned}
\sum_{j=1}^{\infty}\Biggl\|\sum_{i=j}^{\infty} F_i(x|\mathcal{F}_0) - F_i\bigl(x|\mathcal{F}_0^*\bigr)\Biggr\|^2 &\le \sum_{j=1}^{\infty}\Biggl[\sum_{i=j}^{\infty}\bigl\|F_i(x|\mathcal{F}_0) - F_i\bigl(x|\mathcal{F}_0^*\bigr)\bigr\|\Biggr]^2 \\
&\le \sum_{j=1}^{\infty}\Biggl[\sum_{i=j}^{\infty} 2\Biggl\|\min\Biggl\{1, \frac{bC_0}{\sqrt{\omega}}\prod_{k=1}^{i}\bigl(a + b\eta_k^2\bigr)\bigl|\eta_0^2 - \eta_0'^2\bigr|\sigma_0^2\Biggr\}\Biggr\|\Biggr]^2 \\
&\le C\sum_{j=1}^{\infty}\Biggl[\sum_{i=j}^{\infty}\Biggl\|\Biggl(\prod_{k=1}^{i}\bigl(a + b\eta_k^2\bigr)\eta_0^2\sigma_0^2\Biggr)^{\min\{1,\alpha/4\}}\Biggr\|\Biggr]^2 < \infty,
\end{aligned}$$

where the last inequality follows since $\mathrm{E}(a+b\eta_1^2)^{\alpha/4}<1$. Thus, (2.3) of Theorem 2.1 and Theorem 2.2 hold (see also Theorem 2.1 of Chan and Zhang [10]) with

$$a_n=\begin{cases} n^{1/\alpha}, & \text{if } 0<\alpha<2,\\ \sqrt{n\log n}, & \text{if } \alpha=2,\\ \sqrt{n}, & \text{if } \alpha>2. \end{cases}$$

Further, if $f_\eta(\cdot)$ has derivative $f'_\eta(\cdot)$ and $\sup_x f'_\eta(x)<C_0$, then

$$\sum_{j=1}^{\infty}\sup_x\left\|\sum_{i=j}^{\infty}f_i(x|\mathcal{F}_0)-f_i\bigl(x|\mathcal{F}_0^*\bigr)\right\|^2\le C\sum_{j=1}^{\infty}\left[\sum_{i=j}^{\infty}\rho^i\right]^2<\infty.$$

From Theorem 2.1 it follows that for any constant $A>0$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g(S_{i-1}/a_n)\bigl[I(\varepsilon_i<x)-F(x)\bigr]\stackrel{w}{\Longrightarrow}\int_0^1 g\bigl(S(t-)\bigr)\,\mathrm{d}W(t,x),\qquad \text{on } D[-A,A],$$

where $S(t)$ is an α-stable process when $\alpha<2$ and a Gaussian process when $\alpha\ge 2$ and $W(t,x)$ is given as in Theorem 2.1. As a result, when $\beta=1$ in model (1.1),

$$\sup_{x\in[-A,A]}\frac{1}{\sqrt{n}}\alpha_n(x)\stackrel{\mathcal{L}}{\longrightarrow}\sup_{x\in[-A,A]}\int_0^1 g\bigl(S(t)\bigr)\,\mathrm{d}W(t,x).$$

Example 2.2. Let $\{\varepsilon_i\}$ in model (1.1) be an infinite-variance linear moving average process $\varepsilon_i=\sum_{j=1}^{\infty}c_j\eta_{i-j}$, where $\{\eta_i\}$ is an i.i.d. sequence with bounded density and heavy tail index $0<\alpha\le 2$, that is, when $\alpha<2$, $nP(|\eta_1|>a_nx)\to x^{-\alpha}$ for any $x>0$ and when $\alpha=2$, $na_n^{-2}\mathrm{E}(\eta_1^2 I(|\eta_1|\le a_n))\to 1$. Since

$$\begin{aligned}
&\sum_{j=1}^{\infty}\left\|\sum_{i=j}\bigl[F_i(x|\mathcal{F}_0)-F_i\bigl(x|\mathcal{F}_0^*\bigr)\bigr]\right\|^2\\
&\quad\le\sum_{j=1}^{\infty}\left[\sum_{i=j}^{\infty}\bigl\|F_i(x|\mathcal{F}_0)-F_i\bigl(x|\mathcal{F}_0^*\bigr)\bigr\|\right]^2 \qquad (2.12)\\
&\quad\le\sum_{j=1}^{\infty}\left[\sum_{i=j}^{\infty}C\bigl\|\min\bigl(\bigl|c_i\bigl(\eta_0-\eta_0'\bigr)\bigr|,1\bigr)\bigr\|\right]^2\le C'\sum_{j=1}^{\infty}\left[\sum_{i=j}^{\infty}|c_i|^{\alpha/2}\right]^2,
\end{aligned}$$

if $\sum_{j=1}^{\infty}\sum_{i=j}^{\infty}|c_i|^{\alpha/2}<\infty$ (i.e., $\sum_{i=1}^{\infty}i|c_i|^{\alpha/2}<\infty$), condition (i) of (A3) holds. Further, by Chan and Zhang [8] (see also Avram and Taqqu [2]), we also have condition (A1) for $\{\varepsilon_i\}$. Thus, if $\sum_{j=1}^{\infty}\sum_{i=j}^{\infty}|c_i|^{\alpha/2}<\infty$ and condition (A2) holds, then for the unit-root model (1.1), we have for any $x \in \mathbb{R}$,

$$\frac{1}{\sqrt{n}}\alpha_n(x) \xrightarrow{\mathcal{L}} \int_0^1 g\bigl(Z_\alpha(t-)\bigr)\,\mathrm{d}W(t,x), \tag{2.13}$$

where $Z_\alpha(t)$ is a stable process with index α. In particular, when $c_j = j^{-\theta}$ and $\theta > 3/\alpha$, then under condition (A2), conclusion (2.13) holds.

On the other hand, if $0 < \alpha < 1$, since $\sum_{i=1}^n |\varepsilon_i|/a_n = \mathrm{O}_p(1)$, using (ii) of condition (A3), we can relax the condition from $\theta > 3/\alpha$ to $\theta > 2/\alpha$. This observation sheds light on the important subtlety of the roles of θ and α for an infinite variance moving average process.

Example 2.3. When $\{\eta_i\}$ in Example 2.2 has finite variance and bounded density $f_\eta(x)$ and $c_j = j^{-\theta} l(j)$, as $j \to \infty$, for some slowly varying function, similar to (2.12), we have that under $\theta > 1$, Theorem 2.2 holds. Further, if $f_\eta'(x)$ exists and $\sup_x |f_\eta'(x)| \le C_0$ for some $C_0 > 0$ and $\theta > 3/2$, then

$$\begin{aligned}
&\sum_{j=1}^\infty \Biggl\| \sum_{i=j}^\infty f_i(x|\mathcal{F}_0) - f_i\bigl(x|\mathcal{F}_0^*\bigr) \Biggr\|^2 \\
&\quad \le \sum_{j=1}^\infty \Biggl[\sum_{i=j}^\infty C \bigl\| \min\bigl(\bigl|c_i\bigl(\eta_0 - \eta_0'\bigr)\bigr|, 1\bigr) \bigr\| \Biggr]^2 \\
&\quad \le C' \sum_{j=1}^\infty \Biggl[\sum_{i=j}^\infty i^{-\theta} l(i)\Biggr]^2 \le C'' \sum_{j=1}^\infty i^{-2\theta+2} l'(i) < \infty,
\end{aligned}$$

where $l'(x)$ is a slowly varying function, it follows that condition (A4) holds. Thus, if condition (A2) holds,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n g(S_{i-1}/\sqrt{n})\bigl[I(\varepsilon_i < x) - F(x)\bigr] \xrightarrow{\mathcal{L}} \int_0^1 g\bigl(S(t)\bigr)\,\mathrm{d}W(t,x), \qquad \text{on } D[-A,A]$$

and when $\beta = 1$ in model (1.1),

$$\sup_{x\in[-A,A]} \frac{1}{\sqrt{n}} \sum_{i=1}^n g(X_{i-1}/\sqrt{n})\bigl[I(\varepsilon_i < x) - F(x)\bigr] \xrightarrow{\mathcal{L}} \sup_{x\in[-A,A]} \int_0^1 g\bigl(S(t)\bigr)\,\mathrm{d}W(t,x),$$

where $S(t)$ is a Gaussian process.

3. Long-memory error processes

In this section, we study the marked empirical process

$$\alpha_n(x) = \sum_{i=1}^{n} g(X_{i-1}/a_n)\big(I(\varepsilon_i < x) - F(x)\big),$$

when $\{X_t\}$ is a unit root process given by (1.1) with $\beta = 1$ and ε_t being a long-memory process. Long-memory processes have been widely applied in finance and econometrics, see, for example, Baillie [4] and Teyssiére and Kirman [33]. Specifically, let $c_j = j^{-\theta} l(j)$, $l(\cdot)$ be a slowly varying function with $|l(m+n)/l(n) - 1| \le m/n$ for $1 \le m \le n$ and consider the linear moving average process $\varepsilon_i = \sum_{j=1}^{\infty} c_j \eta_{i-j}$ defined in Example 2.3 with $\sum_{j=1}^{\infty} c_j^2 < \infty$ and $\sum_{j=1}^{\infty} |c_j| = \infty$.

The essential idea in studying the weak convergence of $\alpha_n(x)$ when $\{\varepsilon_i\}$ is short-memory is the martingale approximation. This transforms the weak convergence of $\alpha_n(x)$ into those of a martingale stochastic integral $\sum_{i=1}^{n} g(X_{i-1}/a_n)\xi_i(x)$. When $\{\varepsilon_i\}$ is long-memory, this method does not work and the issue of the weak convergence of $\alpha_n(x)$ becomes much more challenging. Fortunately, to circumvent this difficulty, the ideas of Ho and Hsing [20] and Wu [36] become relevant.

Let $f(x)$ be the density of η_1 and $f^{(l)}(x)$ be its lth derivatives and let $f(x) = f^{(0)}(x)$. We have the following results.

Theorem 3.1. *Suppose that $\beta = 1$ in model (1.1) and* (i) $E|\eta_1|^{\nu} < \infty$ *for some* $\nu > \max\{4, 1/(1-\theta)\}$; (ii) $g(\cdot)$ *is Lipschitz on* $\mathbb{R}$; (iii) $\sum_{l=0}^{p} \int_R |f^{(l)}(x)|^2 \,\mathrm{d}x < \infty$, *then if any one of the following three conditions holds*:

(a) $p = 4$ *and* $\theta \in (1/2, 3/4) \cup (5/6, 1)$;
(b) $p = 5$ *and* $\theta \in (1/2, 3/4) \cup (3/4, 1)$ *or* $\theta = 3/4$ *and* $\sum_{i=1}^{\infty} l^4(i)/i < \infty$;
(c) $p = 6$ *and* $\theta \in (1/2, 1)$,

we have

$$\sup_x \frac{1}{a_n} \alpha_n(x) \xrightarrow{\mathcal{L}} \sup_x \int_0^1 g\big(Z_\theta(t)\big) \,\mathrm{d}Z_\theta(t, x). \tag{3.1}$$

If in addition $n(\hat{\beta} - 1) = \mathrm{O}_p(1)$, *then*

$$\begin{aligned} &\sup_x \frac{1}{a_n} \left[\hat{\alpha}_n(x) - f(x)(\hat{\beta} - \beta) \sum_{i=1}^{n} g(X_{i-1}/a_n) X_{i-1} \right] \\ &\quad \xrightarrow{\mathcal{L}} \sup_x f(x) \int_0^1 g\big(Z_\theta(t)\big) \,\mathrm{d}Z_\theta(t), \end{aligned} \tag{3.2}$$

where $a_n = n^{3/2-\theta} l(n)$ *and* $Z_\theta(t) = \int_{-\infty}^{t} \int_0^t [\max(v - u, 0)]^{-\theta} \,\mathrm{d}v \,\mathrm{d}B(u)$, $B(u)$ *is a standard Brownian motion.*

Remark 3.1. When $g(\cdot) \equiv 1$, then Theorem 3.1 reduces to the case of Chan and Ling [7].

4. Proofs

To prove the main results, we need the following lemmas. The first one is due to Lemma 4 of Wu [36].

Lemma 4.1. *Let $H \in \mathcal{C}^1$, the space of functions with continuous first-order derivatives and $a > 0$. Then*

$$\sup_{t \le s \le t+a} H^2(s) \le \frac{2}{a} \int_t^{t+a} H^2(u)\,\mathrm{d}u + 2a \int_t^{t+a} H'^2(u)\,\mathrm{d}u \tag{4.1}$$

and

$$\sup_{t \in \mathbb{R}} H^2(s) \le 2 \int_{\mathbb{R}} H^2(u)\,\mathrm{d}u + 2 \int_{\mathbb{R}} H'^2(u)\,\mathrm{d}u, \tag{4.2}$$

where H' is the derivative of H.

Lemma 4.2. *If $\{\varepsilon_i\}$ is short-memory, then under the conditions* (A1), (A2) *and* (A3), *there exists a martingale difference sequence $\zeta_i(x)$ with respect to $\mathcal{F}_i$ such that for any $\delta > 0$,*

$$\lim_{n\to\infty} P\left\{ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n g(S_{i-1}/a_n)\bigl(I(\varepsilon_i < x) - F(x)\bigr) - \frac{1}{\sqrt{n}} \sum_{i=1}^n g(S_{i-1}/a_n)\zeta_i(x) \right| > \delta \right\} = 0.$$

Proof. When (i) of (A3) holds, then by Volný [34], there exist a random sequence $\xi_i(x) = \sum_{j=-\infty}^{-1} \sum_{l=0}^{\infty} \mathcal{P}_{i+j} I(\varepsilon_{i+l} \le x) \in L^2$ and a martingale difference sequence $\zeta_i(x) = \sum_{j=i}^{\infty} \mathcal{P}_i I(\varepsilon_j \le x)$ such that $I(\varepsilon_i < x) - F(x) = \zeta_i(x) + \xi_i(x) - \xi_{i+1}(x)$. This gives that

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n g(S_{i-1}/a_n)\bigl(I(\varepsilon_i < x) - F(x)\bigr) - \frac{1}{\sqrt{n}} \sum_{i=1}^n g(S_{i-1}/a_n)\zeta_i(x) \\ &\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^{n-1} \xi_{i+1}(x)\bigl[g(S_i/a_n) - g(S_{i-1}/a_n)\bigr] - \frac{1}{\sqrt{n}} g(S_{n-1}/a_n)\xi_{n+1}(x) =: I_1 + I_2. \end{aligned} \tag{4.3}$$

Since for any $\delta > 0$,

$$P\Bigl\{ \sup_{2 \le i \le n+1} \bigl|\xi_i(x)\bigr| > \delta\sqrt{n} \Bigr\} \le \sum_{i=2}^{n+1} (\sqrt{n}\delta)^{-2} \mathrm{E}\bigl[\xi_1^2(x) I\bigl(\bigl|\xi_1(x)\bigr| > \delta\sqrt{n}\bigr)\bigr] \to 0, \tag{4.4}$$

it follows that $|\xi_{n+1}(x)|/\sqrt{n} = \mathrm{o}_p(1)$. On the other hand, by (A1) and (A2), we have $g(S_{n-1}/a_n) = \mathrm{O}_p(1)$. Thus, $I_2 = \mathrm{o}_p(1)$. It suffices to show that $I_1 = \mathrm{o}_p(1)$. When $\{\varepsilon_i\}$ has infinite variance with tail index $\alpha < 2$, the result $I_1 = \mathrm{o}_p(1)$ follows along exactly the lines of argument of Lemma 2 of Knight [25]. We therefore only give the proof for the finite variance case in detail.

When $\{\varepsilon_i\}$ has finite variance or has infinite variance with tail index $\alpha = 2$, since $\sum_{i=0}^{\infty}\theta_2(i) < \infty$, it follows from Theorem 1 of Wu [37] that $\mathrm{E}(\sum_{i=1}^{n}\varepsilon_i)^2 = C_1 n$. Thus, $a_n = C_2\sqrt{n}$. By (A2) and (i) of (A3) for any $\epsilon > 0$, we have

$$\begin{aligned} P(|I_1| > \epsilon) &\le \frac{C}{\sqrt{n}}\sum_{i=1}^{n-1} E\bigl|\xi_{i+1}(x)(\varepsilon_i/a_n)^{\nu} I\bigl(|\varepsilon_i| \le \delta a_n\bigr)\bigr| + P\Bigl(\sup_{1\le i\le n}|\varepsilon_1| > \delta a_n\Bigr) \\ &\le \frac{C'\delta^{\nu-1}}{n}\sum_{i=1}^{n-1}\bigl\{\bigl[\mathrm{E}\xi_{i+1}^2(x)\bigr]^{1/2}\bigl[\mathrm{E}\varepsilon_i^2\bigr]^{1/2}\bigr\} \le \frac{C''\delta^{\nu-1}}{n}\sum_{i=2}^{n}\bigl\{\bigl[\mathrm{E}\xi_i^2(x)\bigr]^{1/2}\bigr\} \qquad (4.5) \\ &\le \frac{C'''\delta^{\nu-1}}{n}\sum_{i=2}^{n}\Biggl\{\sum_{j=-\infty}^{i-1}\mathrm{E}\Biggl[\sum_{l=0}^{\infty}\bigl[F_{i+l}(x|\mathcal{F}_j) - F_{i+l}\bigl(x|\mathcal{F}_j^*\bigr)\bigr]\Biggr]^2\Biggr\}^{1/2} = \mathrm{o}(1) \end{aligned}$$

by taking $\delta \to 0$. This gives that $I_1 = \mathrm{o}_p(1)$ and therefore Lemma 4.2 holds when (i) of (A3) is true.

When (ii) of (A3) holds, by Corollary 1 of Wu [37], we have

$$\sup_{0\le t\le 1}\Biggl|W_n(t,x) - \sum_{i=1}^{[nt]}\zeta_i(x)\Biggr| = \mathrm{o}\bigl(n^{1/2}\bigr), \qquad \text{a.s.} \tag{4.6}$$

where $W_n(t,x) = \sum_{i=1}^{[nt]}(I(\varepsilon_i \le x) - F(x))$. Combining this with (A2) gives

$$\begin{aligned} &\Biggl|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(S_{i-1}/a_n)\bigl(I(\varepsilon_i < x) - F(x)\bigr) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(S_{i-1}/a_n)\zeta_i(x)\Biggr| \\ &\quad = \frac{1}{\sqrt{n}}\Biggl|\sum_{i=1}^{n}\Biggl(W_n(i/n,x) - \sum_{i=1}^{i}\zeta_i(x)\Biggr)\bigl[g(S_i/a_n) - g(S_{i-1}/a_n)\bigr]\Biggr| \qquad (4.7) \\ &\quad \le C\Biggl(\sup_{0\le t\le 1}\frac{1}{\sqrt{n}}\Biggl|W_n(t,x) - \sum_{i=1}^{[nt]}\zeta_i(x)\Biggr|\Biggr)\Biggl(\sum_{i=1}^{n}\frac{|\varepsilon_i|}{a_n}\Biggr) = \mathrm{o}_p(1). \end{aligned}$$

This completes the proof of Lemma 4.2. □

Lemma 4.3. *If* $\{\varepsilon_i\}$ *is short-memory, then under the conditions* (A1), (A2) *and* (A4), *for any constant* $A > 0$,

$$\sup_{x\in[-A,A]}\Biggl|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(S_{i-1}/a_n)\bigl(I(\varepsilon_i < x) - F(x)\bigr) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(S_{i-1}/a_n)\zeta_i(x)\Biggr|$$

converges to zero in probability, where $\zeta_i(x)$ *is defined as in Lemma* 4.2.

Proof. From the proof of Lemma 4.2 for case of (i) of (A3), it suffices to show

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\sup_{x\in[-A,A]}\xi_i^2(x)=\mathrm{O}(1). \tag{4.8}$$

Since $\xi_i(x)=\sum_{j=-\infty}^{i-1}\sum_{l=0}^{\infty}\mathcal{P}_j I(\varepsilon_{i+l}\le x)\in C^1$, it follows from Lemma 4.1 and Fubini's theorem that

$$\begin{aligned}
\mathrm{E}\sup_{x\in[-A,A]}\xi_i^2(x) &\le \frac{2}{A}\int_{-A}^{A}\mathrm{E}\xi_i^2(u)\,\mathrm{d}u+2A\int_{-A}^{A}\mathrm{E}\xi_i'^2(u)\,\mathrm{d}u\\
&\le \frac{2}{A}\int_{-A}^{A}\sum_{j=-\infty}^{i-1}\mathrm{E}\Biggl[\sum_{l=0}^{\infty}\bigl[F_{i+l}(u|\mathcal{F}_j)-F_{i+l}\bigl(u|\mathcal{F}_j^*\bigr)\bigr]\Biggr]^2\mathrm{d}u\\
&\quad+2A\int_{-A}^{A}\sum_{j=-\infty}^{i-1}\mathrm{E}\Biggl[\sum_{l=0}^{\infty}\bigl[f_{i+l}(u|\mathcal{F}_j)-f_{i+l}\bigl(u|\mathcal{F}_j^*\bigr)\bigr]\Biggr]^2\mathrm{d}u\\
&\le 2\sum_{j=-\infty}^{i-1}\sup_u\mathrm{E}\Biggl[\sum_{l=0}^{\infty}\bigl[F_{i+l}(u|\mathcal{F}_j)-F_{i+l}\bigl(u|\mathcal{F}_j^*\bigr)\bigr]\Biggr]^2\\
&\quad+4A^2\sum_{j=-\infty}^{i-1}\max_u\mathrm{E}\Biggl[\sum_{l=0}^{\infty}\bigl[f_{i+l}(u|\mathcal{F}_j)-f_{i+l}\bigl(u|\mathcal{F}_j^*\bigr)\bigr]\Biggr]^2.
\end{aligned}$$

Thus, by (A4), we have (4.8) as desired. □

Lemma 4.4. *Let $\widetilde{W}_n(t,x)=\sum_{i=1}^{[nt]}\zeta_i(x)$, $\zeta_i(x)$ is the martingale difference defined in Lemma 4.2. Then under condition* (A4),

$$\frac{1}{\sqrt{n}}\widetilde{W}_n(t,x)\stackrel{w}{\Longrightarrow}W(t,x)\qquad \textit{on } D\bigl([0,1]\times[-A,A]\bigr).$$

Proof. Condition (A4) implies $\sum_{j=1}^{\infty}\|\sum_{i=j}^{\infty}F_i(x|\mathcal{F}_0)-F_i(x|\mathcal{F}_0^*)\|^2<\infty$, it follows that $E\zeta_i(x)=E\{\sum_{i=0}^{\infty}F_i(x|\mathcal{F}_0)-F_i(x|\mathcal{F}_0^*)\}^2=\mu(x)<\infty$. Since $\{\zeta_i(x)\}$ is a martingale difference sequence, by Theorem 23.1 of Billingsley [6], we have

$$\frac{1}{\sqrt{n}}\widetilde{W}_n(t,x)\stackrel{\mathcal{L}}{\longrightarrow}W(t,x). \tag{4.9}$$

By (4.9) and the Cramér–Wold's device, the finite-dimensional convergence of $\widetilde{W}_n(t,x)$ follows. By Theorem 6 of Bickel and Wichura [5], to show the tightness of $\{\widetilde{W}_n(t,x)\}$ on $D[0,1]\times$

$D[-A, A]$, it suffices to show that for any $0 \le t_1 < t < t_2 \le 1$ and $-A \le x_1 < x < x_2 \le A$,

$$n^{-2}E\left\{\left[\sum_{i=[nt_1]+1}^{[nt]} \zeta_i(x_1, x_2)\right]^2\left[\sum_{i=[nt]+1}^{[nt_2]} \zeta_i(x_1, x_2)\right]^2\right\} \le (t-t_1)(t_2-t)(x_2-x_1)^2 \quad (4.10)$$

and

$$n^{-2}\mathrm{E}\left\{\left|\sum_{i=[nt_1]+1}^{[nt_2]} \zeta_i(x_1, x)\right|^2\left|\sum_{i=[nt_1]+1}^{[nt_2]} \zeta_i(x, x_2)\right|^2\right\} \le C(x-x_1)(x_2-x)(t_2-t_1)^2, \quad (4.11)$$

where $\zeta_i(x, y) = \zeta_i(y) - \zeta_i(x)$. Equations (4.10) and (4.11) follow easily by condition (A4) and noting that $\widetilde{W}_n(t, x)$ is a martingale. Details are omitted. □

Lemma 4.5. *Under the conditions of Theorem 2.1, there exists a dense set $Q \subset [0, 1]$, $0, 1 \in Q$ such that for any finite subset $\{0 \le t_1 < t_2 < \cdots < t_m \le 1\} \subset Q$ and for any x,*

$$\bigl(S_n(t_i)/a_n, \widetilde{W}_n(t_i, x)/\sqrt{n}, 1 \le i \le m\bigr) \xrightarrow{\mathcal{L}} \bigl(S(t_i), W(t_i, x), 1 \le i \le m\bigr). \quad (4.12)$$

Proof. Since $S_n(t) \overset{S}{\Longrightarrow} S(t)$, it follows that there exists a dense set $Q' \subset [0, 1]$, $1 \in Q'$ such that for any finite subset $\{t_1 < t_2 < \cdots < t_m \le 1\} \subset Q'$,

$$a_n^{-1}\bigl(S_n(t_1), S_n(t_2), \ldots, S_n(t_m)\bigr) \xrightarrow{\text{f.d.d.}} \bigl(S(t_1), S(t_2), \ldots, S(t_m)\bigr). \quad (4.13)$$

Note that $S_n(0) = S(0) = 0$. Thus, (4.13) holds for all finite subset of $Q = Q' \cup \{0\}$.

When ε_i has infinite variance, since $W(t, x)$ is a continuous process on $[0, 1] \times [-A, A]$, it follows (see, e.g., page 112 of Billingsley [6]) that the weak convergence of Lemma 4.4 can also be replaced by $C([0, 1] \times C[-A, A])$. Thus, $(S_n(t), \widetilde{W}_n(t, x))$ is uniformly S-tight on $D[0, 1] \times D[0, 1]$. This implies that for any sequence $(S_n(t), \widetilde{W}_n(t, x))$, $t \in Q$, there exists a subsequence $(S_{nk}(t), \widetilde{W}_{nk}(t, x))$ such that

$$\mathbf{Z}_{nk}(t) := \bigl(S_{nk}(t)/a_n, \widetilde{W}_n(t, x)/\sqrt{n}\bigr) \xrightarrow{\mathcal{L}} \mathbf{Z}(t),$$

where $\mathbf{Z}(t)$ is a bivariate random process with marginal distributions $S(t)$ and $W(t, x)$. Following the argument of Theorem 3 in Resnick and Greenwood [31], we have that $S(t)$ and $W(t, x)$ are independent and any convergent subsequence has the same limit. Thus, (4.12) holds.

When ε_i has finite variance, since $\sum_{i=0}^{\infty} \theta_2(i) < \infty$, it follows from Corollary 3 of Dedecker and Merlevède [12] that $S_n(t)/a_n \overset{w}{\Longrightarrow} S(t)$ for some Gaussian process $S(t)$. Thus, by Lemma 4.4, if we can show that for any finite subset $\{t_i, 1 \le i \le m\} \subset [0, 1]$,

$$\bigl(S_n(t_i)/a_n, \widetilde{W}_n(t_i, x)/\sqrt{n}, 1 \le i \le m\bigr) \xrightarrow{\mathcal{L}} \bigl(S(t_i), W(t_i, x), 1 \le i \le m\bigr), \quad (4.14)$$

then $(S_n(t)/a_n, \widetilde{W}_n(t,x)/\sqrt{n}) \overset{w}{\Longrightarrow} (S(t), W(t,x))$ on $D[0,1]$ and (4.12) follows. By Theorem 1 of Wu [37], we have that there exists martingale E_i with respect to $\mathcal{F}_i$ such that

$$\left|\left(S_n(t)/a_n, \widetilde{W}_n(t,x)/\sqrt{n}\right) - \left(\sum_{i=1}^{[nt]} E_i/a_n, \sum_{i=1}^{[nt]} \zeta_i(x)/\sqrt{n}\right)\right| = \mathrm{o}_p(1). \tag{4.15}$$

On the other hand, from the martingale central limit theorem (see Theorem 4.1 of Hall and Heyde [19]), it follows that

$$\sum_{i=1}^{[nt]} \left(E_i/a_n, \zeta_i(x)/\sqrt{n}\right) \overset{w}{\Longrightarrow} \left(S(t), W(t,x)\right). \tag{4.16}$$

Combining (4.15) with (4.16) yields (4.14). This completes the proof of Lemma 4.5. □

Lemma 4.6. *Under the conditions of Theorem* 3.1, *we have*

(a) $\frac{1}{a_n} S_n(t) = \frac{1}{a_n}\sum_{i=1}^{[nt]} \varepsilon_i \overset{w}{\Longrightarrow} Z_\theta(t)$ *on* $D[0,1]$;

(b) $\frac{1}{a_n}\sum_{i=1}^{n} g(S_{i-1}/a_n)\varepsilon_i \xrightarrow{\mathcal{L}} \int_0^1 g(Z_\theta(t))\,\mathrm{d}Z_\theta(t)$.

Proof. (a) can be found in Avram and Taqqu [1]. Next, we give the proof of (b).

With the help of strong approximation, it can be shown that

$$\frac{1}{a_n}\sum_{i=1}^{n} g(S_{i-1}/a_n)\epsilon_i = \frac{1}{a_n}\sum_{i=1}^{n} g(S_{i-1}^*/a_n)\epsilon_i^* + o_p(1),$$

where $S_i^* = \sum_{j=1}^{i} \epsilon_j^*$ and ϵ_j^* is defined similarly to ϵ_j by replacing $\{\eta_i\}$ with i.i.d normal variables $\{\eta_i^*\}$.

Since $Z_\theta(t)$ is a fractional Brownian motion with Hurst index $H = 3/2 - \theta$, by Theorem 4 of Marcus [29], we have

$$P\Big(\limsup_{|s-t|=h\to 0; 0\le s,t\le 1} \big|Z_\theta(s) - Z_\theta(t)\big| < 2h^H \log(1/h)\Big) = 1.$$

Thus, by (a) and $Z_\theta(t)$ is continuous, we have

$$\begin{aligned}
&\lim_{n\to\infty} P\Big\{\limsup_{|s-t|\le h\to 0; 0\le s,t\le 1} \frac{1}{a_n}\big|S_n^*(t) - S_n^*(s)\big| \le 2h^H \log(1/h)\Big\} \\
&\quad = P\Big(\limsup_{|s-t|\le h\to 0; 0\le s,t\le 1} \big|Z_\theta(t) - Z_\theta(s)\big| \le 2h^H \log(1/h)\Big) = 1.
\end{aligned}$$

This implies in probability $S_n^*(t)$ is Hölder continuous with an exponent $a > H$. This gives that, in probability, for any $p > (1/H, \infty)$,

$$\nu_p\big(S_n^*(t)/a_n, [0,1]\big) = \sup_{\kappa} \sum_{i=1}^{m} \big|S_n^*(t_i) - S_n^*(t_{i-1})\big|^p / a_n^p < \infty,$$

where the supremum is taken over all subdivisions κ of $[0,1]: 0 = x_0 < \cdots < x_m = 1, m \geq 1$. Since $g(\cdot)$ is a Lipschitz function, we have in probability

$$\nu_p\bigl(g\bigl(S_n^*(t)/a_n\bigr), [0,1]\bigr) = \sup_\kappa \sum_{i=1}^m \bigl|g\bigl(S_n^*(t_i)/a_n\bigr) - g\bigl(S_n^*(t_{i-1})/a_n\bigr)\bigr|^p < \infty.$$

By the theorem on Stieltjes integrability of Young [38] (see also Theorem 2.4 of Mikosch and Norvaiša [30]), we have in probability the integral

$$\frac{1}{a_n}\sum_{i=1}^n g(S_{i-1}/a_n)\varepsilon_i = \int_0^1 g\bigl(S_n(t-)/a_n\bigr)\,\mathrm{d}S_n(t)/a_n$$

exists. This implies that

$$\begin{aligned}&\int_0^1 g\bigl(S_n(t-)/a_n\bigr)\,\mathrm{d}S_n(t)/a_n\\ &\quad = \lim_{\delta\to 0}\sum_{i=1}^m g\bigl(S_n(t_i)\bigr)\bigl(S_n(t_{i+1}) - S_n(t_i)\bigr)/a_n\end{aligned} \tag{4.17}$$

for some sub-division κ of $[0,1]$: $0 = t_0 < t_1 < \cdots < t_m \leq 1, m = [1/\delta]$ with $t_{i+1} - t_i = \delta$. By (a) and the continuous mapping theorem, we get that for any given m,

$$\begin{aligned}\sum_{i=1}^m g\bigl(S_n(t_i)/a_n\bigr)\bigl(S_n(t_{i+1}) - S_n(t_i)\bigr)/a_n &\xrightarrow{\mathcal{L}} \sum_{i=1}^m g\bigl(Z_\theta(t_i)\bigr)\bigl(Z_\theta(t_{i+1}) - Z_\theta(t_i)\bigr)\\ &\xrightarrow{p} \int_0^1 g\bigl(Z_\theta(t)\bigr)\,\mathrm{d}Z_\theta(t),\end{aligned} \tag{4.18}$$

where the last equality is followed by taking $\delta \to 0$ and the existence of $\int_0^1 g(Z_\theta(t))\,\mathrm{d}Z_\theta(t)$. Combining (4.17) and (4.18) gives (b). The proof of Lemma 4.6 is completed. □

Proof of Theorem 2.1. Lemma 4.5 implies that (5) of Jakubowski [22] holds, that is, there exist a dense set Q such that $(g(S_n(t)), \widetilde{W}_n(t,x)) \xrightarrow{\text{f.d.d.}} (g(S(t)), W(t,x))$. Further, since $g(\cdot)$ is a Lipschitz continuous function and $S_n(t)$ is uniformly S-tight, it follows that $g(S_n(t))$ is also uniformly S-tight. Moreover, for any $x \in \mathbb{R}$, $\widetilde{W}_n(t,x)$ is a martingale satisfying UT condition and is J_1-tight with limiting law concentrated on $C([0,1])$, by Remark 4 of Jakubowski [22], we see that his condition (6) is satisfied. Therefore, for $\{g(S_n(t)), \widetilde{W}_n(t,x)\}$, all the conditions of Theorem 3 of Jakubowski [22] are satisfied, as a result of this theorem, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n g(S_{i-1}/a_n)\zeta_i(x) \xrightarrow{\mathcal{L}} \int_0^1 g\bigl(S(t)\bigr)\,\mathrm{d}W(t,x). \tag{4.19}$$

Thus, (2.3) follows from Lemma 4.2. By Lemma 4.3, for (2.4) it suffices to show that

$$\begin{aligned} U_n(x) &:= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(S_{i-1}/a_n)\zeta_i(x) \\ &\overset{w}{\Longrightarrow} \int_0^1 g\big(S(t)\big)\,\mathrm{d}W(t,x) =: U(x), \qquad \text{on } D[-A,A]. \end{aligned} \tag{4.20}$$

The finite-dimension convergence to (4.20) follows from the Cramér–Wold device and (4.19). Next, we show for any $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$P\Big\{\sup_{|x-y|\le\delta} \big|U_n(x) - U_n(y)\big| > \varepsilon\Big\} \to 0. \tag{4.21}$$

This implies that $U_n(x)$ is tight, as a result, we have (4.20).

Since $S_n(t)/a_n \overset{S}{\Longrightarrow} S(t)$ and $S_n(0) = S(0) = 0$, it follows that

$$\max_{0\le t\le 1} \big|g\big(S_n(t)/a_n\big)\big| \xrightarrow{\mathcal{L}} \max_{0\le t\le 1} \big|g\big(S(t)\big)\big|. \tag{4.22}$$

Let $g_\delta(S_i/a_n) = g(S_i/a_n)I(|g(S_i/a_n)| \le \delta^{-1/4})$ and $V_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^n g_\delta(S_{i-1}/a_n)(\zeta_i(x) - \zeta_i(y))$. Then $V_n(x)$ is a martingale and by Lemma 4.1 and condition (A4),

$$\begin{aligned} &\mathrm{E}\Big[\sup_{y\le x\le y+\delta} \big|V_n(x)\big|\Big]^2 \\ &\quad\le \frac{2}{\delta n}\int_y^{y+\delta} \mathrm{E}\bigg(\sum_{i=1}^n \bigg[g_\delta\bigg(\frac{S_{i-1}}{a_n}\bigg)\big(\zeta_i(u) - \zeta_i(y)\big)\bigg]\bigg)^2 \mathrm{d}u \\ &\qquad + \frac{2\delta}{n}\int_y^{y+\delta} \mathrm{E}\bigg(\sum_{i=1}^n \bigg[g_\delta\bigg(\frac{S_{i-1}}{a_n}\bigg)\zeta_i'(u)\bigg]\bigg)^2 \mathrm{d}u \\ &\quad\le \frac{2\delta^{-1/2}}{n}\sum_{i=1}^n \int_y^{y+\delta} \mathrm{E}\big\{\zeta_i(u) - \zeta_i(y)\big\}^2 \mathrm{d}u + \frac{2\delta^{-1/2}\delta^2}{n}\sum_{i=1}^n \sup_{y\le x\le y+\delta} \mathrm{E}\big\{\zeta_i'(x)\big\}^2 \\ &\quad\le \frac{2\delta^{-1/2}}{n}\sum_{i=1}^n \int_y^{y+\delta}\int_y^u \mathrm{E}\big\{\zeta_i'(a)\big\}^2 \mathrm{d}a\,\mathrm{d}u + \frac{2\delta^{-1/2}\delta^2}{n} \sup_{x\in[-A,A]} \mathrm{E}\big\{\zeta_i'(x)\big\}^2 \le C\delta^{3/2}. \end{aligned} \tag{4.23}$$

Note that

$$\begin{aligned} &P\Big\{\sup_{|x-y|\le\delta} \big|U_n(x) - U_n(y)\big| > 4\varepsilon\Big\} \\ &\quad\le C\big(1 + [A/\delta]\big)P\Big\{\sup_{y\le x\le y+\delta} \big|V_n(x)\big| > \varepsilon\Big\} + P\Big\{\max_{1\le i\le n} \big|g(S_i/a_n)\big| > \delta^{-1/4}\Big\}. \end{aligned}$$

By (4.22), (4.23) and taking δ small enough, we have (4.21) as desired. This completes the proof of Theorem 2.1. □

Proof of Theorem 2.2. Note that when $\beta = 1$, $X_i = X_0 + \sum_{j=1}^{i} \varepsilon_j$ and $X_0/a_n \xrightarrow{p} 0$, (2.5) follows directly from (2.3). Next, we show (2.6).

Let $\{u_{ni}\}$ be a constant sequence with $\max_i |u_{ni}| = \mathrm{o}(1)$. Along the lines of proof in Lemma 4.2, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g\left(\frac{S_{i-1}}{a_n}\right)\bigl[I(\varepsilon_i \le x + u_{ni}) - I(\varepsilon_i \le x)\bigr] \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g\left(\frac{S_{i-1}}{a_n}\right)\bigl[\zeta_i(x + u_{ni}) - \zeta_i(x)\bigr] \\
&\qquad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g\left(\frac{S_{i-1}}{a_n}\right)\bigl(F(x + u_{ni}) - F(x)\bigr) + \mathrm{o}_p(1) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g\left(\frac{S_{i-1}}{a_n}\right)\bigl(F(x + u_{ni}) - F(x)\bigr) + \mathrm{o}_p(1).
\end{aligned} \tag{4.24}$$

Since $\max_{1\le i\le n} |X_i/a_n| = \mathrm{O}_p(1)$, it follows that when $a_n(\hat\beta - \beta) = \mathrm{o}_p(1)$, $\max_{1\le i\le n}(\hat\beta - \beta)X_i = \mathrm{o}_p(1)$. Thus, by (4.24), we have

$$\frac{1}{\sqrt{n}}\bigl(\hat\alpha_n(x) - \alpha_n(x)\bigr) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(S_{i-1}/a_n)\bigl(F\bigl(x + (\hat\beta - \beta)X_{i-1}\bigr) - F(x)\bigr) + \mathrm{o}_p(1).$$

This gives (2.6) and completes the proof of Theorem 2.2. □

Proof of Theorem 2.3. Since $X_i = X_0 + \sum_{j=1}^{i} \varepsilon_j$ and $X_0/a_n \xrightarrow{p} 0$, (2.7) follows from (2.4) and the continuous mapping theorem. Let u_{ni} be given as that in the proof of (4.24), then by Lemma 4.3 and a similar argument of (4.21), we have that under the condition (A4),

$$\begin{aligned}
&\sup_{x\in[-A,A]} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_{i-1}/a_n)\bigl[I(\varepsilon_i \le x + u_{ni})\bigr] \\
&\quad = \sup_{x\in[-A,A]} \Biggl[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_{i-1}/a_n) I(\varepsilon_i \le x) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_{i-1}/a_n)\bigl[\zeta_i(x + u_{ni}) - \zeta_i(x)\bigr] \\
&\qquad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_{i-1}/a_n)\bigl(F(x + u_{ni}) - F(x)\bigr)\Biggr] + \mathrm{o}_p(1) \\
&\quad = \sup_{x\in[-A,A]} \Biggl[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_{i-1}/a_n) I(\varepsilon_i \le x) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_{i-1}/a_n)\bigl(F(x + u_{ni}) - F(x)\bigr)\Biggr] \\
&\qquad + \mathrm{o}_p(1).
\end{aligned}$$

As a result, by $\max_{1\le i\le n}(\hat\beta-1)X_i=o_p(1)$ and Taylor's expansion, we have in probability,

$$\sup_{x\in[-A,A]}\frac{\hat\alpha_n(x)}{\sqrt{n}}=\sup_{x\in[-A,A]}\left[\frac{\alpha_n(x)}{\sqrt{n}}+\frac{1}{\sqrt{n}}f(x)(\hat\beta-1)\sum_{i=1}^{n}g(X_{i-1}/a_n)X_{i-1}\right]. \tag{4.25}$$

Further, by Theorem 3 of Jakubowski [22], it follows that

$$\frac{1}{n}\sum_{i=1}^{n}\bigl[g(X_{i-1}/a_n)X_{i-1}/a_n\bigr]\xrightarrow{w}\int_0^1 g\bigl(S(t)\bigr)S(t)\,\mathrm{d}t. \tag{4.26}$$

Combining equations (2.7), (4.25), (4.25) with Corollary 2.1 yields (2.8). Equations (2.9) and (2.10) follow similarly by noting that when $\hat\beta$ is the LSE of β, then

$$\begin{aligned}
n(\hat\beta-\beta) &= \frac{1}{2}\left[X_n^2/a_n^2-X_0^2/a_n^2-\sum_{i=1}^{n}\varepsilon_i^2/a_n^2\right]\Big/\left[\frac{1}{n}\sum_{i=1}^{n}X_{i-1}^2/a_n^2\right]\\
&\xrightarrow{w}\frac{1}{2}\bigl(S^2(1)-S^2\bigr)\Big/\int_0^1 S^2(t)\,\mathrm{d}t\\
&=:\int_0^1 S(t-)\,\mathrm{d}S(t)\Big/\int_0^1 S^2(t)\,\mathrm{d}t.
\end{aligned}$$

The proof of Theorem 2.3 is completed. □

Proof of Theorem 3.1. Since the proof of the three cases are similar, we only give the proof under condition (b) in details. Let $U_{l,i}=\sum_{0\le j_1<\cdots<j_i}\prod_{s=1}^{i}c_{j_s}\eta_{l-j_s}$, $U_{l,0}=1$ and $L(\widetilde\varepsilon_l,x,k)=I(\varepsilon_l\le x)-\sum_{i=0}^{k}(-i)^iF^{(i)}(x)U_{l,i}$. By Lemma 10 of Wu [36], we have that for all x,

$$\begin{aligned}
\bigl\|\mathcal{P}_1\bigl(L(\widetilde\varepsilon_i,x,3)\bigr)\bigr\| &= \mathrm{O}\left\{|c_{i-1}|\left[|c_{i-1}|+\left(\sum_{j=i}^{\infty}|c_j|^4\right)^{1/2}+\left(\sum_{j=i}^{\infty}|c_j|^2\right)^{1/2}\right]\right\}\\
&=\mathrm{O}\bigl(i^{-2\theta}l^2(i)+i^{-4\theta+3/2}l^3(i)\bigr).
\end{aligned} \tag{4.27}$$

Thus, when $\theta>3/4$ or $\theta=3/4$ and $\sum_{i=1}^{\infty}l^4(i)/i<\infty$, for all x,

$$\begin{aligned}
\sum_{j=1}^{\infty}\left\|\sum_{i=j}^{\infty}\mathcal{P}_1\bigl(L(\widetilde\varepsilon_i,x,3)\bigr)\right\|^2 &\le \sum_{j=1}^{\infty}\left(\sum_{i=j}^{\infty}\bigl\|\mathcal{P}_1\bigl(L(\widetilde\varepsilon_i,x,3)\bigr)\bigr\|\right)^2\\
&=\mathrm{O}\left[\sum_{j=1}^{\infty}\left(\sum_{i=j}^{\infty}i^{-2\theta}l^2(i)\right)^2\right]<\infty.
\end{aligned}$$

By Theorem 2 of Volný [34], there exists a martingale difference sequence $D_i(x) \in L^2$ and a finite variance sequence $\{e_i(x)\}$ such that for all x,

$$L(\widetilde{\varepsilon}_i, x, 3) = D_i(x) + e_i(x) - e_{i+1}(x).$$

Applying (ii) instead of (i) of Lemma 4.1 in proving Lemma 4.3, we have that

$$\sup_{x\in\mathbb{R}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(S_{i-1}/a_n)L(\widetilde{\varepsilon}_i, x, 3) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(S_{i-1}/a_n)D_i(x)\right| = \mathrm{o}_p(1).$$

Let $g_M(x) = g(x)I(|g(x)| \le M)$. By (ii) of Lemma 4.1, we have

$$\begin{aligned}
&\mathrm{E}\left(\sup_{x\in\mathbb{R}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_M(S_{i-1}/a_n)D_i(x)\right)^2\\
&\quad\le \frac{2}{n}\mathrm{E}\int_{\mathbb{R}}\left(\sum_{i=1}^{n} g_M(S_{i-1}/a_n)D_i(x)\right)^2 \mathrm{d}x + \frac{2}{n}\mathrm{E}\int_{\mathbb{R}}\left(\sum_{i=1}^{n} g_M(S_{i-1}/a_n)D_i'(x)\right)^2 \mathrm{d}x = \mathrm{O}\bigl(M^2\bigr).
\end{aligned}$$

As a results, for any positive constants ε and η, there exist a large M_0 and a large N_0 such that for all $M > M_0$ and $n > N_0$,

$$\begin{aligned}
&P\left\{\sup_{x\in\mathbb{R}}\frac{1}{a_n}\left|\sum_{i=1}^{n} g\left(\frac{S_{i-1}}{a_n}\right)L(\widetilde{\varepsilon}_i, x, 3)\right| > 2\varepsilon\right\}\\
&\quad\le P\left\{\sup_{x\in\mathbb{R}}\frac{1}{a_n}\left|\sum_{i=1}^{n} g_M\left(\frac{S_{i-1}}{a_n}\right)D_i(x)\right| > \varepsilon\right\} + P\left\{\max_{1\le i\le n}\left|g\left(\frac{S_i}{a_n}\right)\right| > M\right\} + \eta \le 3\eta.
\end{aligned} \tag{4.28}$$

Note that

$$\begin{aligned}
\frac{1}{a_n}\sum_{i=1}^{n} g\left(\frac{S_{i-1}}{a_n}\right)U_{i,2} &= \left(\frac{n^{2-2\theta}l^2(n)}{a_n}\right)g\left(\frac{S_{n-1}}{a_n}\right)\sum_{i=1}^{n}\frac{U_{i,2}}{n^{1-2(\theta-1/2)}}\\
&\quad - \frac{n^{2-2\theta}l^2(n)}{a_n}\sum_{i=1}^{n-1}\sum_{j=1}^{i}\frac{U_{i,2}}{n^{2-2\theta}l^2(n)}\left[g\left(\frac{S_i}{a_n}\right) - g\left(\frac{S_{i-1}}{a_n}\right)\right].
\end{aligned} \tag{4.29}$$

From Avram and Taqqu [1], it follows that there exists a constant $C(\theta)$ such that

$$\begin{aligned}
\sum_{i=1}^{[nt]}\frac{U_{i,2}}{n^{2-2\theta}l^2(n)} &\xrightarrow{\mathcal{L}} C(\theta)\int_{\mathbb{R}}\int_{\mathbb{R}}\int_0^t\prod_{i=1}^{2}\bigl[\max(0, v-u_i)\bigr]^{-\theta}\,\mathrm{d}v\,\mathrm{d}B(u_1)\,\mathrm{d}B(u_2)\\
&=: Z_{2,\theta}(t).
\end{aligned} \tag{4.30}$$

By (4.30), the Lipschitz condition of $g(\cdot)$ and an argument similar to Theorem 3.1 of Ling and Li [28], we have that the right-hand side of (4.29) converges to zero in probability. Further, by Lemma 4.1, we have

$$\sup_{x\in\mathbb{R}}\bigl(f^{(k)}(x)\bigr)^2 \le 2\int_{\mathbb{R}}\bigl(f^{(k)}(x)\bigr)^2\,\mathrm{d}x + 2\int_{\mathbb{R}}\bigl(f^{(k+1)}(x)\bigr)^2\,\mathrm{d}x < \infty \qquad \text{for all } k \le p-1.$$

Thus, by (4.28), we have $\sup_{x\in\mathbb{R}}\frac{1}{a_n}\sum_{i=1}^n g(S_{i-1}/a_n)[I(\varepsilon_i \le x) - F(x) + f(x)\varepsilon_i] = \mathrm{o}_p(1)$. Combining this with Lemma 4.6 gives that

$$\sup_{x\in\mathbb{R}}\frac{1}{a_n}\sum_{i=1}^n g(S_{i-1}/a_n)\bigl[I(\varepsilon_i \le x) - F(x)\bigr] \xrightarrow{\mathcal{L}} \sup_{x\in\mathbb{R}} f(x)\int_0^1 g\bigl(Z_\theta(t)\bigr)\,\mathrm{d}Z_\theta(t). \tag{4.31}$$

When $1/2 < \theta < 3/4$, by (ii) of Theorem 3 in Wu [36], we have

$$\frac{1}{n^{2-2\theta}l^2(n)}\sum_{i=1}^n\bigl[I(\varepsilon_i \le x) - F(x) + f(x)\varepsilon_i\bigr] \stackrel{w}{\Longrightarrow} f'(x)Z_{2,\theta}(1), \qquad \text{on } D(\mathbb{R}). \tag{4.32}$$

Using (4.32), we also have (4.31). By noting that as $\{X_t\}$ is a unit root process, $X_t = S_t + X_0$. This completes the proof of (3.1).

Applying (4.31) and arguing as in Theorem 2.3, we have that when $n(\hat\beta - 1) = \mathrm{O}_p(1)$,

$$\sup_{x\in\mathbb{R}}\left[\frac{1}{a_n}\bigl(\hat\alpha_n(x) - \alpha_n(x)\bigr) - \frac{1}{a_n}\sum_{i=1}^n g(X_{i-1}/a_n)\bigl[F\bigl(x + (\hat\beta - 1)X_{i-1}\bigr) - F(x)\bigr]\right] = \mathrm{o}_p(1).$$

Since $\sup_{x\in\mathbb{R}} f(x) < \infty$ and $\sup_{1\le i\le n}(\hat\beta - 1)X_i = \mathrm{O}_p(a_n/n)$, it follows from Taylor's expansion and (3.1) that

$$\sup_{x\in\mathbb{R}}\frac{1}{a_n}\left[\hat\alpha_n(x) - f(x)(\hat\beta - 1)\sum_{i=1}^n g(X_{i-1}/a_n)X_{i-1}\right] \xrightarrow{\mathcal{L}} \sup_{x\in\mathbb{R}} f(x)\int_0^1 g\bigl(Z_\theta(t)\bigr)\,\mathrm{d}Z_\theta(t).$$

This gives (3.2) and completes the proof of Theorem 3.1. □

Acknowledgements

We would like to thank the Editor, the Associate Editor and two anonymous referees for helpful comments, which led to an improved version of this paper. This research was supported in part by grants from the General Research Fund of HKSAR-RGC-GRF Nos. 400408 and 400410, Collaborative Research Fund of HKSAR-RGC-CRF: CityU8/CRF/09, Fundamental Research Funds for the Central Universities, ZJNSF (No. R6090034) and NSFC (Nos. 11171074 and 10801118). Part of this research was completed when the second author (the corresponding author) visited CUHK in 2011–2012. Research support from the Statistics Department of CUHK is gratefully acknowledged.

References

[1] Avram, F. and Taqqu, M.S. (1987). Noncentral limit theorems and Appell polynomials. *Ann*. *Probab*. **15** 767–775. MR0885142

[2] Avram, F. and Taqqu, M.S. (1992). Weak convergence of sums of moving averages in the α-stable domain of attraction. *Ann*. *Probab*. **20** 483–503. MR1143432

[3] Bai, J.S. (2003). Testing parametric conditional distributions of dynamic models. *The Review of Economics and Statistics* **85** 531–549.

[4] Baillie, R.T. (1996). Long memory processes and fractional integration in econometrics. *J*. *Econometrics* **73** 5–59. MR1410000

[5] Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann*. *Math*. *Statist*. **42** 1656–1670. MR0383482

[6] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396

[7] Chan, N.H. and Ling, S. (2008). Residual empirical processes for long and short memory time series. *Ann*. *Statist*. **36** 2453–2470. MR2458194

[8] Chan, N.H. and Zhang, R.M. (2009). Inference for nearly nonstationary processes under strong dependence with infinite variance. *Statist*. *Sinica* **19** 925–947. MR2536137

[9] Chan, N.H. and Zhang, R.M. (2009). Quantile inference for near-integrated autoregressive time series under infinite variance and strong dependence. *Stochastic Process*. *Appl*. **119** 4124–4148. MR2565561

[10] Chan, N.H. and Zhang, R.M. (2010). Inference for unit-root models with infinite variance GARCH errors. *Statist*. *Sinica* **20** 1363–1393. MR2777329

[11] Davis, R. and Resnick, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann*. *Statist*. **14** 533–558. MR0840513

[12] Dedecker, J. and Merlevède, F. (2003). The conditional central limit theorem in Hilbert spaces. *Stochastic Process*. *Appl*. **108** 229–262. MR2019054

[13] del Barrio, E., Deheuvels, P. and van de Geer, S. (2007). *Lectures on Empirical Processes*: *Theory and Statistical Applications*. *EMS Series of Lectures in Mathematics*. Zürich: Eur. Math. Soc. With a preface by Juan A. Cuesta Albertos and Carlos Matrán. MR2284824

[14] Escanciano, J.C. (2006). Goodness-of-fit tests for linear and nonlinear time series models. *J*. *Amer*. *Statist*. *Assoc*. **101** 531–541. MR2256173

[15] Escanciano, J.C. (2007). Weak convergence of non-stationary multivariate marked processes with applications to martingale testing. *J*. *Multivariate Anal*. **98** 1321–1336. MR2364121

[16] Escanciano, J.C. (2010). Asymptotic distribution-free diagnostic tests for heteroskedastic time series models. *Econometric Theory* **26** 744–773. MR2646478

[17] Gaenssler, P. and Révész, P. (1976). *Empirical Distributions and Processes*. *Lecture Notes in Mathematics* **566**. Berlin: Springer. MR0428364

[18] Gnedenko, B.V. and Kolmogorov, A.N. (1954). *Limit Distributions for Sums of Independent Random Variables*. Cambridge, MA: Addison-Wesley. Translated and annotated by K.L. Chung. With an Appendix by J.L. Doob. MR0062975

[19] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application* (*Probability and Mathematical Statistics*). New York: Academic Press [Harcourt Brace Jovanovich Publishers]. MR0624435

[20] Ho, H.C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann*. *Statist*. **24** 992–1024. MR1401834

[21] Hong, Y. and Lee, T.H. (2003). Diagnostic checking for the adequacy of nonlinear time series models. *Econometric Theory* **19** 1065–1121. MR2015977

[22] Jakubowski, A. (1996). Convergence in various topologies for stochastic integrals driven by semimartingales. *Ann*. *Probab*. **24** 2141–2153. MR1415245

[23] Jakubowski, A. (1997). A non-Skorohod topology on the Skorohod space. *Electron. J. Probab.* **2** 21 (electronic). MR1475862
[24] Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131** 207–248. MR0440724
[25] Knight, K. (1991). Limit theory for M-estimates in an integrated infinite variance process. *Econometric Theory* **7** 200–212. MR1128412
[26] Koul, H.L. and Ling, S. (2006). Fitting an error distribution in some heteroscedastic time series models. *Ann. Statist.* **34** 994–1012. MR2283401
[27] Kurtz, T.G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19** 1035–1070. MR1112406
[28] Ling, S. and Li, W.K. (1998). Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscedastic errors. *Ann. Statist.* **26** 84–125. MR1611800
[29] Marcus, M.B. (1968). Hölder conditions for Gaussian processes with stationary increments. *Trans. Amer. Math. Soc.* **134** 29–52. MR0230368
[30] Mikosch, T. and Norvaiša, R. (2000). Stochastic integral equations without probability. *Bernoulli* **6** 401–434. MR1762553
[31] Resnick, S. and Greenwood, P. (1979). A bivariate stable characterization and domains of attraction. *J. Multivariate Anal.* **9** 206–221. MR0538402
[32] Stute, W., Xu, W.L. and Zhu, L.X. (2008). Model diagnosis for parametric regression in high-dimensional spaces. *Biometrika* **95** 451–467. MR2521592
[33] Teyssière, G. and Kirman, A. P. (2007). *Long Memory in Economics*. Berlin: Springer. MR2263582
[34] Volný, D. (1993). Approximating martingales and the central limit theorem for strictly stationary processes. *Stochastic Process. Appl.* **44** 41–74. MR1198662
[35] Wooldridge, J.M. (1990). A unified approach to robust, regression-based specification tests. *Econometric Theory* **6** 17–43. MR1059144
[36] Wu, W.B. (2003). Empirical processes of long-memory sequences. *Bernoulli* **9** 809–831. MR2047687
[37] Wu, W.B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389
[38] Young, L.C. (1936). An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.* **67** 251–282. MR1555421
[39] Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.* **36** 1108–1126. MR2418651

Received September 2011 and revised March 2012